

OIPE

RAW SEQUENCE LISTING DATE: 10/04/2001
PATENT APPLICATION: US/09/866,307 TIME: 17:40:34

Input Set : A:\sequence.txt
Output Set: N:\CRF3\10042001\I866307.raw

SEQUENCE LISTING

3 (1) GENERAL INFORMATION:

- 4 (i) APPLICANT: LAWRENCE P. WACKETT
- 5 (ii) TITLE OF INVENTION: DNA MOLECULES AND PROTEIN DISPLAYING
- 6 IMPROVED TRIAZINE COMPOUND DEGRADING ABILITY
- 7 (iii) NUMBER OF SEQUENCES: 26
- 8 (iv) CORRESPONDENCE ADDRESS:
- 9 (A) ADDRESSEE: MUETING, RAASCH & GEBHARDT, P.A.
- 10 (B) STREET: 119 North Fourth Street
- 11 (C) CITY: Minneapolis
- 12 (D) STATE: Minnesota
- 13 (E) COUNTRY: USA
- 14 (F) ZIP: 55401

ENTERED

15 (v) COMPUTER READABLE FORM:

- 16 (A) MEDIUM TYPE: Floppy disk
- 17 (B) COMPUTER: IBM PC compatible
- 18 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
- 19 (D) SOFTWARE: PatentIn Release #1.0, Version #1.30

20 (vi) CURRENT APPLICATION DATA:

- C--> 21 (A) APPLICATION NUMBER: US/09/866,307
- C--> 22 (B) FILING DATE: 20-Sep-2001
- 23 (C) CLASSIFICATION:

24 (vii) PRIOR APPLICATION DATA:

- 25 (A) APPLICATION NUMBER: 60/035,404
- 26 (B) FILING DATE: 17-JAN-1997
- 27 (A) APPLICATION NUMBER: PCT/US98/00944
- 28 (B) FILING DATE: 16-JAN-1998
- 29 (A) APPLICATION NUMBER: 09/155,036
- 30 (B) FILING DATE: 17-SEPT-1998

31 (viii) ATTORNEY/AGENT INFORMATION:

- 32 (A) NAME: PROVENCE, DAVID L.
- 33 (B) REGISTRATION NUMBER: 43,022
- 34 (C) REFERENCE/DOCKET NUMBER: 110.00440102

35 (ix) TELECOMMUNICATION INFORMATION:

- 36 (A) TELEPHONE: 612-305-1005
- 37 (B) TELEFAX: 612-305-1228

39 (2) INFORMATION FOR SEQ ID NO: 1:

41 (i) SEQUENCE CHARACTERISTICS:

- 42 (A) LENGTH: 1858 base pairs
- 43 (B) TYPE: nucleic acid
- 44 (C) STRANDEDNESS: single
- 45 (D) TOPOLOGY: linear

47 (ii) MOLECULE TYPE: DNA (genomic)

50 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 1:

52 CTCGGGTAAC TTCTTGAGCG CGGCCACAGC AGCCTTGATC ATGAAGGCGA GCATGGTGAC	60
54 CTTGACGCCG CTCTTTTCGT TCTCTTTGTT GAACTGCACG CGAAAGGCTT CCAGGTCGGT	120
56 GATGTCCGCG TCGTCGTGGT TGGTGACGTG CGGGATGACC ACCCAGTTGC GGTGCAGGTT	180

RAW SEQUENCE LISTING

DATE: 10/04/2001

PATENT APPLICATION: US/09/866,307

TIME: 17:40:34

Input Set : A:\sequence.txt

Output Set: N:\CRF3\10042001\I866307.raw

```

58 TTTTCGATGGC ATAATATCTG CGTTGCGACG TGTAACACAC TATTGGAGAC ATATCATGCA      240
60 AACGCTCAGC ATCCAGCAGC GTACCCTCGT CACGATGGAT CAGTACCGCA GAGTCCTTGG      300
62 GGATAGCTGG GTTACAGTGC AGGATGGACG GATCGTCGCG CTCGGAGTGC ACGCCGAGTC      360
64 GGTGCTCCG CCAGCGGATC GGGTGATCGA TGCACGCGGC AAGGTCGTGT TACCCGGTTT      420
66 CATCAATGCC CACACCCATG TGAACCAGAT CCTCCTGCGC GGAGGGCCCT CGCACGGACG      480
68 TCAATTCAT GACTGGCTGT TCAACGTTGT GTATCCGGA CAAAAGGCGA TGAGACCGGA      540
70 GGACGTAGCG GTGGCGGTGA GGTGTATTG TGCGGAAGCT GTGCGCAGCG GGATTACGAC      600
72 GATCAACGAA AACGCCGATT CGGCCATCTA CCCAGGCAAC ATCGAGGCCG CGATGGCGGT      660
74 CTATGGTGAG GTGGGTGTGA GGTGCGTCTA CGCCCGCATG TTCTTTGATC GGATGGACGG      720
76 GCGCATTCAA GGGTATGTGG ACGCCTTGAA GGCTCGCTCT CCCCAAGTCG AACTGTGCTC      780
78 GATCATGGAG GAAACGGCTG TGGCCAAAGA TCGGATCACA GCCCTGTCAG ATCAGTATCA      840
80 TGGCACGGCA GGAGGTCGTA TATCAGTTTG GCCCGCTCCT GCCACTACCA CGGCGGTGAC      900
82 AAGTTGAAGGA ATGGCGACGA TCAAGCCTT GCCTCGTGAT CGGGCGGTAA TGTGGACGCT      960
84 TCACATGGCG GAGAGCGATC ATGATGAGCG GATTTCATGGG ATGAGTCCCG CCGAGTACAT     1020
86 GGAGTGTTAC GGACTCTTGG ATGAGCGTCT GCAGGTCGCG CATTGCGTGT ACTTTGACCG     1080
88 GAAGGATGTT CGGCTGCTGC ACCGCCACAA TGTGAAGGTC GCGTCGCAGG TTGTGAGCAA     1140
90 TGCTACCTC GGCTCAGGGG TGGCCCCCGT GCCAGAGATG GTGGAGCGCG GCATGGCCGT     1200
92 GGGCATTGGA ACAGATAACG GGAATAGTAA TGACTCCGCA AACATGATCG GAGACATGAA     1260
94 GTTTATGGCC CATATTCACC GCGCGGTGCA TCGGGATGCG GACGTGCTGA CCCAGAGAA     1320
96 GATTCTTGAA ATGGCGACGA TCGATGGGGC CGCTTCGTTG GGAATGGACC ACGAGATTGG     1380
98 TTCCATCGAA ACCGGCAAGC GCGCGGACCT TATCTGCTT GACCTGCGTC ACCTCAGACC     1440
100 ACTCTCACAT CATTGGCGG CCACGATCGT GTTTCAGGCT TACGGCAATG AGGTGGACAC     1500
102 TGCTCTGATT GACGGAAACG TTGTGATGGA GAACCGCCG TTAGGCTTTC TTCCCCCTGA     1560
104 ACGTGAGTTG GCGTTCCTTG AGGAAGCGCA GAGCCGCGCC ACAGCTATTT TGACGCGGGC     1620
106 GAACATGGTG GCTAACCCAG CTGCGGCAG CCTCTAGGAA ATGACGCCGT TGCTGCATCC     1680
108 GCCGCCCCCT GAGGAAATCG CTGCCATCTT GCGCGGCTC GGATTGGGG GCGGACATGA     1740
110 CCTTGATGGA TACGAATTC CATGAATGC GGCCTTCGCT TCCTTCGCTC GTGTGGAATC     1800
112 GTTGGTAGGT GAGGGTCGAC TCGGGCGGCC AGCTTCCCGA AGAGGTGAAA GGCCCGAG     1858

```

114 (2) INFORMATION FOR SEQ ID NO: 2:

116 (i) SEQUENCE CHARACTERISTICS:

117 (A) LENGTH: 473 amino acids

118 (B) TYPE: amino acid

119 (C) STRANDEDNESS: single

120 (D) TOPOLOGY: linear

122 (ii) MOLECULE TYPE: protein

127 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 2:

```

129 Met Gln Thr Leu Ser Ile Gln His Gly Thr Leu Val Thr Met Asp Gln
130 1 5 10 15
132 Tyr Arg Arg Val Leu Gly Asp Ser Trp Val His Val Gln Asp Gly Arg
133 20 25 30
135 Ile Val Ala Leu Gly Val His Ala Glu Ser Val Pro Pro Ala Asp
136 35 40 45
138 Arg Val Ile Asp Ala Arg Gly Lys Val Val Leu Pro Gly Phe Ile Asn
139 50 55 60
141 Ala His Thr His Val Asn Gln Ile Leu Leu Arg Gly Gly Pro Ser His
142 65 70 75 80
144 Gly Arg Gln Phe Tyr Asp Trp Leu Phe Asn Val Val Tyr Pro Gly Gln
145 85 90 95
147 Lys Ala Met Arg Pro Glu Asp Val Ala Val Ala Val Arg Leu Tyr Cys

```

RAW SEQUENCE LISTING

PATENT APPLICATION: US/09/866,307

DATE: 10/04/2001

TIME: 17:40:34

Input Set : A:\sequence.txt

Output Set: N:\CRF3\10042001\I866307.raw

```

148          100          105          110
150 Ala Glu Ala Val Arg Ser Gly Ile Thr Thr Ile Asn Glu Asn Ala Asp
151          115          120          125
153 Ser Ala Ile Tyr Pro Gly Asn Ile Glu Ala Ala Met Ala Val Tyr Gly
154          130          135          140
156 Glu Val Gly Val Arg Val Val Tyr Ala Arg Met Phe Phe Asp Arg Met
157          145          150          155          160
159 Asp Gly Arg Ile Gln Gly Tyr Val Asp Ala Leu Lys Ala Arg Ser Pro
160          165          170          175
162 Gln Val Glu Leu Cys Ser Ile Met Glu Glu Thr Ala Val Ala Lys Asp
163          180          185          190
165 Arg Ile Thr Ala Leu Ser Asp Gln Tyr His Gly Thr Ala Gly Gly Arg
166          195          200          205
168 Ile Ser Val Trp Pro Ala Pro Ala Thr Thr Thr Ala Val Thr Val Glu
169          210          215          220
171 Gly Met Arg Trp Ala Gln Ala Phe Ala Arg Asp Arg Ala Val Met Trp
172          225          230          235          240
174 Thr Leu His Met Ala Glu Ser Asp His Asp Glu Arg Ile His Gly Met
175          245          250          255
177 Ser Pro Ala Glu Tyr Met Glu Cys Tyr Gly Leu Leu Asp Glu Arg Leu
178          260          265          270
180 Gln Val Ala His Cys Val Tyr Phe Asp Arg Lys Asp Val Arg Leu Leu
181          275          280          285
183 His Arg His Asn Val Lys Val Ala Ser Gln Val Val Ser Asn Ala Tyr
184          290          295          300
186 Leu Gly Ser Gly Val Ala Pro Val Pro Glu Met Val Glu Arg Gly Met
187          305          310          315          320
189 Ala Val Gly Ile Gly Thr Asp Asn Gly Asn Ser Asn Asp Ser Ala Asn
190          325          330          335
192 Met Ile Gly Asp Met Lys Phe Met Ala His Ile His Arg Ala Val His
193          340          345          350
195 Arg Asp Ala Asp Val Leu Thr Pro Glu Lys Ile Leu Glu Met Ala Thr
196          355          360          365
198 Ile Asp Gly Ala Arg Ser Leu Gly Met Asp His Glu Ile Gly Ser Ile
199          370          375          380
201 Glu Thr Gly Lys Arg Ala Asp Leu Ile Leu Leu Asp Leu Arg His Leu
202          385          390          395          400
204 Arg Arg Leu Ser His His Leu Ala Ala Thr Ile Val Phe Gln Ala Tyr
205          405          410          415
207 Gly Asn Glu Val Asp Thr Val Leu Ile Asp Gly Asn Val Val Met Glu
208          420          425          430
210 Asn Arg Arg Leu Ser Phe Leu Pro Pro Glu Arg Glu Leu Ala Phe Leu
211          435          440          445
213 Glu Glu Ala Gln Ser Arg Ala Thr Ala Ile Leu Gln Arg Ala Asn Met
214          450          455          460
216 Val Ala Asn Pro Ala Trp Arg Ser Leu
217          465          470
219 (2) INFORMATION FOR SEQ ID NO: 3:
221 (i) SEQUENCE CHARACTERISTICS:

```

RAW SEQUENCE LISTING

DATE: 10/04/2001

PATENT APPLICATION: US/09/866,307

TIME: 17:40:34

Input Set : A:\sequence.txt

Output Set: N:\CRF3\10042001\I866307.raw

```

222      (A) LENGTH: 1808 base pairs
223      (B) TYPE: nucleic acid
224      (C) STRANDEDNESS: single
225      (D) TOPOLOGY: linear
227      (ii) MOLECULE TYPE: DNA (genomic)
232      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 3:
234 GCGAGCATGG TGACCTTGAC GCCGCTCTTT TCGTTCTCTT TGTTGAACTG CACGCGAAAG      60
236 GCTTCCAGGT CCGTGATGTC CGCGTCGTCG TGTTTGGTGA CGTGCGGGAT GACCACCCAG      120
238 TTGCGGTGCA GGTTTTTCGA TGGCATAATA TCTGCGTTGC GACGTGTAAC AACTATTG      180
240 AGACATATCA TGCAAACGCT CAGCATCCAG CACGGTACCC TCGTCACGAT GGATCAGTAC      240
242 CGCAGAGTCC TTGGGGATAG CTGGGTTTAC GTGCAGGATG GACGGATCGT CGCGCTCGGA      300
244 GTGCACGCCG AGTCGGTGCC TCCGCCAGCG GATCGGGTGA TCGATGCACG CGGCAAGGTC      360
246 GTGTTACCCG GTTTCATCAA TGCCACACCC CATGTGAACC AGATCCTCCT GCGCGGAGGG      420
248 CCCTCGCAGC GCGCTCAATT CTATGACTGG CTGTTCAACG TTGTGTATCC GGGACAAAAG      480
250 GCGATGAGAC CGGAGGACGT AGCGGTGGCG GTGAGGTTGT ATTGTGCGGA AGCTGTGCGC      540
252 AGCGGGATTA CGACGATCAA CGAAAACGCC GATTCGGCCA TCTACCCAGG CAACATCGAG      600
254 GCCCGCATGG CGGTCTATGG TGAGGTGGGT GTGAGGGTCG TCTACGCCCG CATGTTCTTT      660
256 GATCGGATGG ACGGGCGCAT TCAAGGGTAT GTGGACGCCT TGAAGGCTCG CTCTCCCCAA      720
258 GTCGAACGTG GCTCGATCAT GGAGGGAACG GACTGGGCCA AAGATCGGAT CACAGCCCTG      780
260 TCAGATCAGT ATCATGGCAC GGCAGGAGGT CGTATATCAG TTTGGCCCGC TCCTGCCACT      840
262 ACCACGGCGG TGACAGTTGA AGGAATGCGA TGGGCACAAG CCTTCGCCCG TGATCGGGCG      900
264 GTAATGTGGA CGCTTCACAT GCGGAGAGC GATCATGATG AGCGGATTCA TGGGATGAGT      960
266 CCCGCCGAGT ACATGGAGTG TTACGGACTC TTGGATGAGC GTCTGCAGGT CGCGCATTCG      1020
268 GTGTACTTTG ACCGGAAGGA TGTTCCGGCTG CTGCACCGCC ACAATGTGAA GGTGCGGTCG      1080
270 CAGGTTGTGA GCAATGCCTA CCTCGGCTCA GGGGTGGCCC CCGTGCCAGA GATGGTGGAG      1140
272 CGCGGCATGG CCGTGGGCAT TGGAACAGAT AACGGGAATA GTAATGACTC CGTAAACATG      1200
274 ATCGGAGACA TGAAGTTTAT GGCCCATATT CACCGCGCGG TGCATCGGGA TCGGACGTG      1260
276 CTGACCCAGC AGAAGATTCT TGAAATGGCG ACGATCGATG GGGCGCGTTC GTTGGGAATG      1320
278 GACCACGAGA TTGGTTCCAT CGAAACCGGC AAGCGCGCGG ACCTTATCCT GCTTGACCTG      1380
280 CGTCACCCTC AGACGACTCC TCACCATCAT TTGGCGGCCA CGATCGGTGTT TCAGGCTTAC      1440
282 GGCAATGAGG TGGACACTGT CCTGATTGAC GGAACCGTTG TGATGGAGAA CCGCCGCTTG      1500
284 AGCTTTCTTC CCCCTGAACG TGAGTTGGCG TTCCTTGAGG AAGCGCAGAG CCGCGCCACA      1560
286 GCTATTTTGC AGCGGGCGAA CATGGTGGCT AACCAGCTT GCGCGAGCCT CTAGGAAATG      1620
288 ACGCCGTTGC TGCAATCCGCC GCCCCTTGAG GAAATCGCTG CCATCTTGGC GCGGCTCGGA      1680
290 TTGGGGGGCG GACATGACCT TGATGGATAC AGAATTGCCA TGAATGCGGC ACTTCCGTCC      1740
292 TTCGCTCGTG TGGAATCGTT GGTAGGTGAG GGTGCGACTG GGGCGCCAGC TTCCCGAAGA      1800
294 AGTGAAAG
296 (2) INFORMATION FOR SEQ ID NO: 4:
298      (i) SEQUENCE CHARACTERISTICS:
299          (A) LENGTH: 1846 base pairs
300          (B) TYPE: nucleic acid
301          (C) STRANDEDNESS: single
302          (D) TOPOLOGY: linear
304      (ii) MOLECULE TYPE: DNA (genomic)
309      (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 4:
311 GAGCGCCGCC ACAGCAGCCT TGATCATGAA GCGCAGCATG GTGACCTTGA CGCCGCTCTT      60
313 TTCGTTCTCT TTGTGAACT GCACGCGAAA GGCTTCCAGG TCGGTGATGT CCGGTCGTC      120
315 GTGGTTGGTG ACGTCCGGA TGACCACCA GTTGCGGTGC AGGTTTTTCG ATGGCGTAAT      180
317 ATCTGCGTTG CGACGTGTAA CACACTATTG GAGACATATC ATGCAAACGC TCAGCATCCA      240

```

RAW SEQUENCE LISTING

DATE: 10/04/2001

PATENT APPLICATION: US/09/866,307

TIME: 17:40:34

Input Set : A:\sequence.txt

Output Set: N:\CRF3\10042001\I866307.raw

```

319 GCACGGTACC CTCGTCACGA TGGATCAGTA CCGCAGAGTC CTTGGGGATA GCTGGGTTCA 300
321 CGTGCAGGAT GGACGGATCG TCGCGCTCGG AGTGCACGCC GAGTCGGTGC CTCCGCCAGC 360
323 GGATCGGGTG ATCGATGCAC GCGGCAAGGT CGTGTTACCC GGTTCATCA ATGCCACAC 420
325 CCATGTGAAC CAGATCCTCC TCGCGGAGG GCCCTCGCAC GGGCGTCAAT TCTATGACTG 480
327 GCTGTTCAAC GTTGTGTATC CGGGACAAAA GGCGATGAGA CCGGAGGACG TAGCGGTGGC 540
329 GGTGAGGTTG TATTGTGCGG AAGCTGTGCG CAGCGGGATT ACGACGATCA ACGAAAACGC 600
331 CGATTGCGCC ATCTACCCAG GCAACATCGA GGCCGCGATG GCGGTCTATG GTGAGGTGGG 660
333 TGTGAGGGTC GTCTACGCCC GCATGTTCTT TGATCGGATG GACGGGCGCA TTCAAGGGTA 720
335 TGTGGACGCC TTGAAGGCTC GCTCTCCCA AGTCGAACTG TGCTCGATCA TGGAGGAAAC 780
337 GGCTGTGGCC AAAGATCGGA TCACAGCCCT GTCAGATCAG TATCATGGCA CGGCAGGAGG 840
339 TCGTATATCA GTTTGGCCCG CTCCTGCCAC TACCACGGCG GTGACAGTTG AAGGAATGCG 900
341 ATGGGCACAA GCCTTCGCCG GTGATCGGGC GGTAAATGTG ACGCTTCACA TGGCGGAGAG 960
343 CGATCATGAT GAGCGGATTC ATGGGATGAG TCCCGCCGAT TACATGGAGT GTTACGGACT 1020
345 CTTGGATGAG CGTCTGCAGG TCGCGCATTG CGTGTACTTT GACCGGAAGG ATGTTCCGGCT 1080
347 GCTGCACCGC CACAATGTGA AGGTCGCGTC GCAGGTTGTG AGCAATGCCT ACCTCGGCTC 1140
349 AGGGGTGGCC CCCGTGCCAG AGATGGTGGA GCGCGGCATG GCCGTGGGCA TTGGAACAGA 1200
351 TAACGGGAAT AGTAATGACT CCGTAAACAT GATCGGAGAC ATGAAGTTTA TGGCCCATAT 1260
353 TCACCGCGCG GTGCATCGGG ATGCGGACGT GCTGACCCCA GAGAAGATTG TTGAAATGGC 1320
355 GACGATCGAT GGGGCGCGTT CGTTGGGGAT GGACCACGAG ATTGGTTCCA TCGAAACCGG 1380
357 CAAGCGCGCG GACCTTATCC TGCTTGACCT GCGTCACCT CAGACGACTC CTCACCATCA 1440
359 TTTGGCGGCC ACGATCGTGT TTCAGGCTTA CGGCAATGAG GTGGACACTG TCCTGATTGA 1500
361 CGGAAACGTT GTGATGGAGA ACCGCGCTT GAGCTTCTT CCCCCTGAAC GTGAGTTGGC 1560
363 GTTCCTTGAG GAAGCGCAGA GCCGCGCCAC AGCTATTTTG CAGCGGGCGA ACATGTTGGC 1620
365 TAACCCAGCT TGGCGCAGCC TCTAGGAAAT GACGCCGTTG CTGCATCCGC CGCCCCTTGA 1680
367 GGAAATCGCT GCCATCTTGG CGCGGCTCGG ATTGGGGGGC GGACATGACC TTGATGGATA 1740
369 CAGAATTGCC ATGAATGCGG CACTTCCGTC CTTGCTCGT GTGGAATCGT TGGTAGGTGA 1800
371 GGTGCGACTG CGGGCGCCAG CTTCCGAAG AAGTGAAAGG CCCGAG 1846

```

373 (2) INFORMATION FOR SEQ ID NO: 5:

375 (i) SEQUENCE CHARACTERISTICS:

376 (A) LENGTH: 601 amino acids

377 (B) TYPE: amino acid

378 (C) STRANDEDNESS: single

379 (D) TOPOLOGY: linear

381 (ii) MOLECULE TYPE: protein

386 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 5:

```

388 Ala Ser Met Val Thr Leu Thr Pro Leu Phe Ser Phe Ser Leu Leu Asn
389 1 5 10 15
391 Cys Thr Arg Lys Ala Ser Arg Ser Val Met Ser Ala Ser Ser Trp Leu
392 20 25 30
394 Val Thr Cys Gly Met Thr Thr Gln Leu Arg Cys Arg Phe Phe Asp Gly
395 35 40 45
397 Ile Ile Ser Ala Leu Arg Arg Val Thr His Tyr Trp Arg His Ile Met
398 50 55 60
400 Gln Thr Leu Ser Ile Gln His Gly Thr Leu Val Thr Met Asp Gln Tyr
401 65 70 75 80
403 Arg Arg Val Leu Gly Asp Ser Trp Val His Val Gln Asp Gly Arg Ile
404 85 90 95
406 Val Ala Leu Gly Val His Ala Glu Ser Val Pro Pro Pro Ala Asp Arg
407 100 105 110

```

VERIFICATION SUMMARY

PATENT APPLICATION: US/09/866,307

DATE: 10/04/2001

TIME: 17:40:35

Input Set : A:\sequence.txt

Output Set: N:\CRF3\10042001\I866307.raw

L:21 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]

L:22 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]